

Unit-I: Infrastructure and Platforms for Building Applications using AI

Part-1: Hardware used in building AI applications: Processors - CPU, GPU, TPU, NPU, Memory - RAM, VRAM, Storage - HDD, SSD

Part-2: Platforms for building applications using AI: Online platforms (Example - Google AutoML, H2O.ai, Teachable Machine or similar platforms - for practice only); Desktop (No-code/Lowcode) platforms (Orange Data Mining, KNIME, Weka, RapidMiner or similar tools - for practice only).

Part-3: Edge AI: Concept; Applications in daily life in devices like Refrigerators, Led Bulbs, Surveillance Cameras, Micro Ovens, Smart Cars/Scooters; Edge AI in smart Appliances

Part-1: Hardware used in Building AI applications

❖ Processors Used in Building AI Applications:

Artificial Intelligence applications require powerful processors to perform large numbers of computations such as matrix operations, neural network training, and data analysis. AI models process huge datasets and therefore need hardware capable of handling complex mathematical calculations efficiently. Different types of processors are used depending on the type of AI task, performance requirements, and energy efficiency. The most commonly used processors in AI systems include CPU (Central Processing Unit), GPU (Graphics Processing Unit), TPU (Tensor Processing Unit), and NPU (Neural Processing Unit).

1. CPU (Central Processing Unit):

The CPU (Central Processing Unit) is the main processor of a computer that executes instructions and controls the overall operation of the system. It is often called the brain of the computer because it processes data and coordinates the activities of all hardware components.

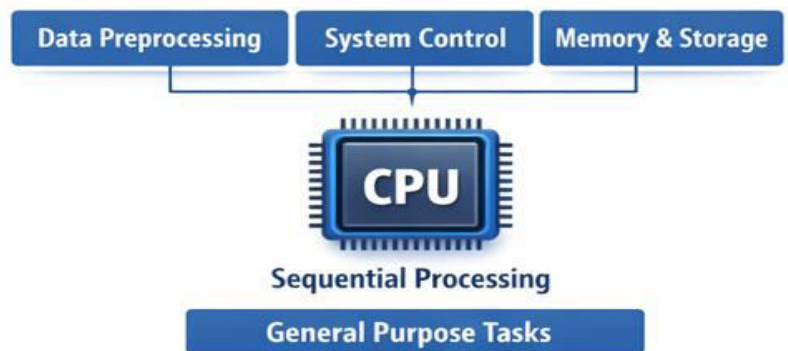


Fig (1): Block Diagram of CPU Architecture for AI Applications

✚ Characteristics:

- ✓ Designed mainly for sequential processing.
- ✓ Contains a small number of powerful cores for complex computations.
- ✓ Highly flexible and capable of performing multiple types of tasks.
- ✓ Suitable for general computing and system-level operations.

✚ Role in AI Applications:

- ✓ Controls the overall functioning of the AI system.
- ✓ Performs data preprocessing tasks such as normalization, filtering, and dataset loading.
- ✓ Manages communication between hardware components like RAM, storage devices, and GPUs.
- ✓ Executes AI programs and machine learning algorithms written in languages like Python.
- ✓ Coordinates the workflow between software frameworks and hardware accelerators.

✚ Example Uses in AI:

- ✓ Running Python AI scripts and machine learning programs.
- ✓ Managing AI pipelines in Jupyter Notebook or IDEs.
- ✓ Performing data preparation before model training.
- ✓ Controlling AI frameworks such as TensorFlow and PyTorch.

2. GPU (Graphics Processing Unit):

The GPU (Graphics Processing Unit) is a specialized processor designed to handle large numbers of calculations simultaneously. It was originally developed for rendering graphics and images, but today it is widely used in Artificial Intelligence and deep learning applications because of its ability to perform parallel computations.

✚ Characteristics:

- ✓ Contains thousands of small cores for parallel processing.
- ✓ Provides much faster computation than CPUs for large datasets.
- ✓ Highly efficient for deep learning model training.
- ✓ Widely used in high-performance computing and AI research.

✚ Role in AI Applications:

- ✓ Performs large-scale parallel computations required in machine learning and deep learning.
- ✓ Accelerates training of neural networks by processing multiple calculations simultaneously.
- ✓ Handles matrix and vector operations used in AI algorithms.
- ✓ Supports AI frameworks such as TensorFlow and PyTorch for faster model training.
- ✓ Used in applications like computer vision, speech recognition, and natural language processing.

✚ Example Uses in AI:

- ✓ Training deep learning models.
- ✓ Image recognition and object detection systems.
- ✓ Autonomous vehicle vision systems.
- ✓ Video analysis and speech processing.

3. TPU (Tensor Processing Unit):

The TPU (Tensor Processing Unit) is a specialized processor developed by Google specifically for machine learning and deep learning tasks. It is designed to accelerate tensor computations, which are the core mathematical operations used in neural networks.

✚ Characteristics:

- ✓ Designed specifically for machine learning workloads.
- ✓ Provides very high computational speed for tensor operations.
- ✓ More energy efficient compared to traditional processors for AI tasks.
- ✓ Mostly used in cloud computing environments.

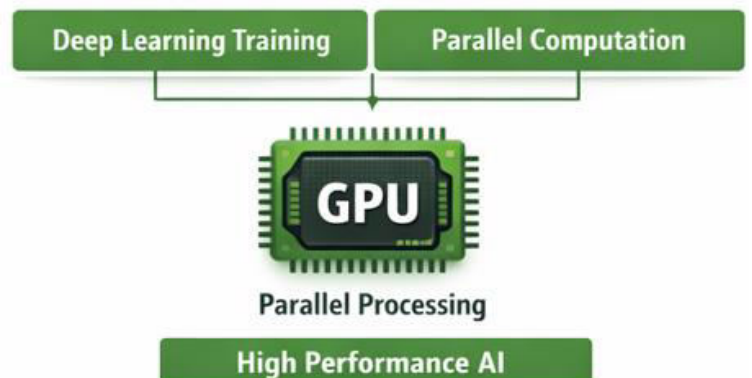


Fig (2): Block Diagram of GPU Architecture for AI Applications

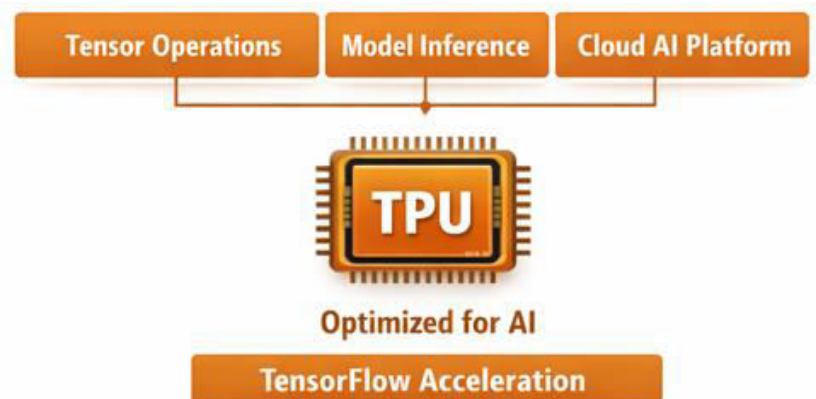


Fig (3): Block Diagram of TPU Architecture for AI Applications

✚ **Role in AI Applications:**

- ✓ Performs tensor operations such as matrix multiplications and transformations used in neural networks.
- ✓ Accelerates training and inference of deep learning models.
- ✓ Used in cloud platforms for large-scale AI model development.
- ✓ Supports machine learning frameworks such as TensorFlow.
- ✓ Enables faster processing of large AI datasets.

✚ **Example Uses in AI:**

- ✓ Image classification systems.
- ✓ Language translation models.
- ✓ Recommendation systems used by online platforms.
- ✓ Large-scale AI research and data center applications.

4. NPU (Neural Processing Unit):

The NPU (Neural Processing Unit) is a processor specifically designed to accelerate artificial intelligence operations, particularly neural network inference tasks in edge devices. NPUs are commonly integrated into smartphones, IoT devices, and smart appliances.

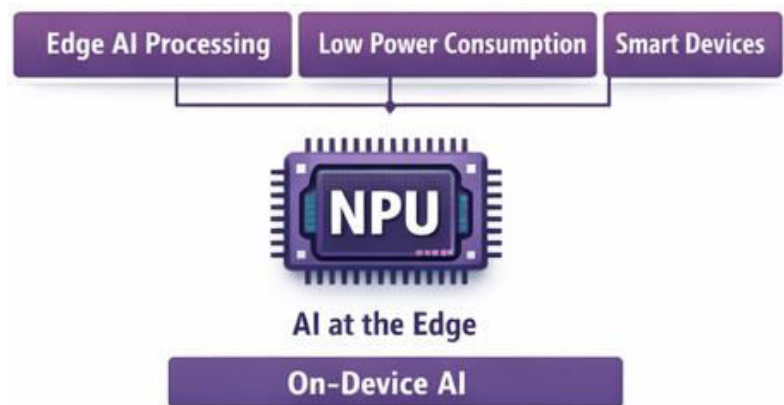


Fig (4): Block Diagram of NPU Architecture for AI Applications

✚ **Characteristics:**

- ✓ Optimized for deep learning inference operations.
- ✓ Consumes very low power, making it suitable for mobile and IoT devices.
- ✓ Provides fast real-time AI processing.
- ✓ Integrated into modern smart devices and embedded systems.

✚ **Role in AI Applications:**

- ✓ Executes neural network computations directly on devices.
- ✓ Enables Edge AI, where AI processing happens locally instead of the cloud.
- ✓ Supports real-time AI applications such as face recognition and voice assistants.
- ✓ Reduces latency and dependence on internet connectivity.
- ✓ Improves data privacy by processing information locally.

✚ **Example Uses in AI:**

- ✓ Smartphone AI cameras and face recognition.
- ✓ Smart surveillance cameras.
- ✓ Voice assistants in smart speakers.
- ✓ Smart home appliances and wearable devices.

❖ **Memory:**

- ✓ Memory is a computer component used to store data, instructions, and results required for processing. It allows the CPU and other processors to access information quickly while executing programs.
- ✓ In Artificial Intelligence (AI) systems, memory is used to store datasets, program instructions, model parameters, and intermediate results during processing. This helps AI algorithms perform computations efficiently.

1. RAM (Random Access Memory):

RAM (Random Access Memory) is the primary memory of a computer used to temporarily store data and instructions that are currently being processed by the CPU. It provides fast read and write access to the processor, enabling efficient execution of programs.

✚ Key Characteristics:

- ✓ **Volatile Memory:** Data is lost when power is turned OFF.
- ✓ **High Speed:** Much faster than storage devices like HDD and SSD.
- ✓ **Temporary Storage:** Holds data only while processing is happening.
- ✓ **Directly Accessible by CPU:** Enables fast execution of programs.
- ✓ **Supports Multitasking:** Allows multiple programs to run simultaneously.

✚ Role of RAM in AI Applications: In Artificial Intelligence systems, RAM plays an important role in:

- ✓ **Loading AI Programs:** AI libraries such as TensorFlow and PyTorch are loaded into RAM for execution.
- ✓ **Handling Input Data:** Datasets used for machine learning are temporarily stored in RAM before computation.
- ✓ **Preprocessing Operations:** Tasks like normalization, resizing images, and tokenizing text occur in RAM.
- ✓ **Model Execution (Inference):** During prediction, data moves from RAM to the CPU/GPU for processing, producing the output results.
- ✓ **Training Machine Learning Models:** Intermediate calculations and model parameters are temporarily stored in RAM.
- ✓ **Multitasking in AI Development:** Running Jupyter Notebook, IDEs, and dataset loaders simultaneously requires large RAM capacity.

✚ Example: If you train a small machine learning model on student data:

1. Dataset loads into RAM
2. CPU/GPU performs calculations using the data
3. Intermediate results are stored temporarily in RAM

2. VRAM (Video Random Access Memory):

VRAM (Video Random Access Memory) is a specialized type of memory located inside the GPU (Graphics Processing Unit) used to store data required for high-speed parallel computation, especially in AI, graphics rendering, and deep learning tasks.

✚ Key Characteristics:

- ✓ Located on the Graphics Card (GPU) rather than on the motherboard.
- ✓ Designed for massive parallel processing required by GPUs.
- ✓ Provides higher memory bandwidth than normal RAM.
- ✓ Stores tensors, matrices, feature maps, and model weights during AI computation.
- ✓ Optimized for high-speed data transfer between GPU cores.

✚ Role of VRAM in AI Applications: VRAM is extremely important for deep learning and GPU-based AI processing.



Fig (1): RAM

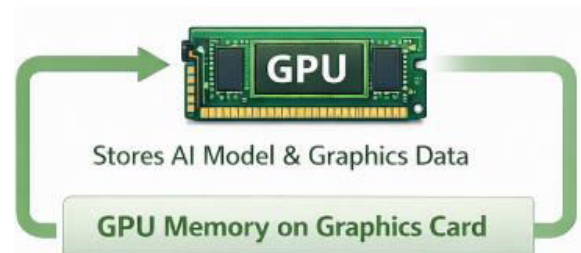


Fig (2): VRAM

- ✓ **Stores Neural Network Models:** Entire AI models including weights and layers are loaded into VRAM.
- ✓ **Handles Large Matrix Operations:** Deep learning involves heavy matrix multiplications, which are executed using GPU and VRAM.
- ✓ **Batch Processing of Data:** Training datasets such as images, audio, or text batches are stored in VRAM during processing.
- ✓ **Accelerates Training Speed:** Larger VRAM allows larger batch sizes, which speeds up training.
- ✓ **Supports Parallel AI Processing:** Thousands of GPU cores access VRAM simultaneously for high-speed computation.

✚ **Example:** When training a Convolutional Neural Network (CNN) for image recognition:

1. Images loaded into VRAM
2. GPU performs convolution operations
3. Gradients stored in VRAM
4. Backpropagation executed at high speed.

❖ **Storage Devices:**

- ✓ Storage devices are used to permanently store data, software, operating systems, and AI datasets in a computer system. Unlike RAM, storage is non-volatile, meaning the data remains even when the power is turned off.
- ✓ In AI applications, storage is used to keep large datasets, trained models, and experimental results. The two common storage types are Hard Disk Drive (HDD) and Solid State Drive (SSD). HDD provides large capacity at low cost, while SSD provides faster data access and better performance.

1. **HDD (Hard Disk Drive):**

HDD (Hard Disk Drive) is a traditional non-volatile storage device that stores data using magnetic recording on rotating disks called platters. It is used for permanent storage of operating systems, software, datasets, and backups.

✚ **Characteristics of HDD:**

- ✓ **Non-Volatile:** Data remains even after power is turned OFF.
- ✓ **Large Storage Capacity:** Available in terabytes (TB) at relatively low cost.
- ✓ **Mechanical Device:** Contains moving parts such as spinning platters and read/write heads.
- ✓ Slower Data Access compared to SSDs.
- ✓ Cost-Effective for storing large volumes of data

✚ **Role of HDD in AI Infrastructure:** HDD is mainly used where large storage capacity is more important than speed.

- ✓ **Archival of Large Datasets:** Stores large raw datasets such as images, videos, and sensor data.
- ✓ **Backup Storage:** Used for saving trained AI models and research data.
- ✓ **Cold Storage in AI Labs:** Data that is not frequently accessed is stored on HDD.
- ✓ **Institutional Data Repositories:** Universities and research institutions store years of experimental data economically.
- ✓ **Dataset Libraries for Training:** Large AI datasets are often stored in HDD before being loaded into faster memory for processing.

❖ **Example in AI Workflow:** Dataset is initially stored in the HDD, then it is moved to SSD or RAM for processing. After that, AI model training is performed, and finally the results and trained models are saved back to the HDD for long-term storage.

2. SSD (Solid State Drive):

SSD (Solid State Drive) is a non-volatile storage device that permanently stores data using flash memory chips instead of rotating disks (as in HDD). It provides very fast data access, which makes it highly suitable for modern computing and AI systems.

✚ Key Characteristics:

- ✓ **Non-Volatile Memory:** Data is retained even when power is OFF.
- ✓ **No Moving Parts:** More reliable and durable than HDD.
- ✓ **Very High Speed:** Much faster read/write operations.
- ✓ **Low Latency:** Quick loading of datasets and software.
- ✓ **Energy Efficient:** Consumes less power.
- ✓ **Silent Operation:** No mechanical noise because there are no moving parts.

✚ Role of SSD in AI Applications: SSD plays a crucial role in delivering data quickly to RAM and VRAM during AI workflows.

- ✓ **Dataset Storage:** Large AI datasets such as images, videos, and text corpora are stored on SSD.
- ✓ **Faster Data Loading:** High read speed reduces AI model training time.
- ✓ **Model Storage:** Trained models (such as .pt, .h5, and .onnx files) can be saved and retrieved quickly.
- ✓ **Fast Data Pipeline:** AI training requires continuous reading of data; SSD prevents data bottlenecks.
- ✓ **Operating System and AI Tools:** Software such as Python, TensorFlow, PyTorch, and MATLAB run faster when installed on SSD.
- ✓ **Supports Big Data Processing:** Important for handling terabytes of machine learning datasets.

✚ Example in AI Workflow: During image classification training:

- ✓ Images are stored in SSD
- ✓ Data is loaded quickly into RAM
- ✓ Sent to VRAM (GPU) for training
- ✓ Results and trained models are saved back to SSD
- ✓

Part-2: Platforms for Building Applications using AI

❖ Online Platforms for Building AI Applications

- ✓ Online platforms are cloud-based tools that allow users to build, train, and deploy Artificial Intelligence (AI) models through a web browser without requiring complex hardware or advanced programming skills. They provide ready-made machine learning tools, automated model training, and data processing features, making AI development easier for beginners, students, and researchers.
- ✓ These platforms typically offer graphical interfaces, automated machine learning (AutoML), and pre-built algorithms. Users can upload datasets, train models, test predictions, and deploy AI applications online without installing heavy software or maintaining high-performance computing systems.

✚ Key Features of Online AI Platforms

- ✓ **Cloud-based access:** AI models can be developed using a web browser without installing software.
- ✓ **Automated Machine Learning (AutoML):** Automatically selects algorithms, tunes parameters, and trains models.
- ✓ **User-friendly interface:** Many platforms provide drag-and-drop tools suitable for beginners.

- ✓ **Scalable computing resources:** Cloud servers provide powerful processors and GPUs for model training.
- ✓ **Easy model deployment:** Trained models can be integrated into applications through APIs.

✚ **Advantages:**

- ✓ No need for high-performance local hardware.
- ✓ Easy to use for students and beginners.
- ✓ Supports collaborative development and remote access.
- ✓ Provides pre-trained models and built-in datasets.

✚ **Examples of Online AI Platforms**

- ✓ **Google AutoML:** A cloud-based platform that allows users to create custom machine learning models automatically without deep programming knowledge.
- ✓ **H2O.ai:** An open-source AI platform that provides automated machine learning, predictive analytics, and scalable AI model development.
- ✓ **Teachable Machine:** A beginner-friendly web tool developed by Google that allows users to train simple AI models for image, sound, and pose recognition directly in a browser.

❖ **Desktop Platforms (No-Code / Low-Code) for Building AI Applications:**

- ✓ Desktop AI platforms are software tools installed on a local computer that allow users to build machine learning and data mining models without extensive programming. These platforms provide graphical user interfaces (GUI), drag-and-drop workflows, and built-in algorithms for developing AI applications.
- ✓ Unlike online platforms, desktop tools run locally on the user's computer, allowing users to work with datasets offline and experiment with machine learning models without internet dependency. They are widely used in education, research, and data analysis for understanding machine learning concepts and building prototype AI models.

❖ **Key Features of Desktop AI Platforms**

- ✓ Graphical user interface (GUI) for easy model development.
- ✓ Drag-and-drop workflow design for building AI pipelines.
- ✓ Provides built-in machine learning algorithms for classification, regression, and clustering.
- ✓ Allows data visualization and model evaluation.
- ✓ Can work offline on local datasets.

❖ **Advantages:**

- ✓ Easy to use for students and beginners.
- ✓ No need for advanced programming knowledge.
- ✓ Suitable for experimental learning and academic projects.
- ✓ Allows visual understanding of machine learning workflows.

❖ **Examples of Desktop AI Platforms:**

- ✓ **Orange Data Mining:** A visual programming tool used for data analysis, machine learning, and data visualization using drag-and-drop components.
- ✓ **KNIME:** A powerful analytics platform that allows users to create data processing and machine learning workflows using a graphical interface.
- ✓ **Weka:** A popular open-source machine learning software that provides classification, clustering, regression, and visualization tools.
- ✓ **RapidMiner:** A data science platform that enables users to design machine learning workflows and predictive models through a visual interface.

Part-3: Edge AI

❖ Concept of Edge AI:

- ✓ Edge AI refers to the implementation of Artificial Intelligence algorithms directly on local devices (edge devices) instead of sending all data to cloud servers for processing.
- ✓ In traditional AI systems, data collected from devices is sent to cloud data centers for analysis, which may cause latency, high bandwidth usage, and dependence on internet connectivity.
- ✓ In contrast, Edge AI performs AI computations locally on the device itself, using embedded processors, microcontrollers, or AI chips. This enables faster decision-making and real-time processing without heavy reliance on the cloud.

✚ Key Characteristics of Edge AI

- ✓ **Local Data Processing:** Data is processed on the device itself instead of being sent to remote servers.
- ✓ **Low Latency:** Immediate response because processing occurs near the data source.
- ✓ **Reduced Internet Dependency:** Devices can function even with limited or no internet connection.
- ✓ **Improved Privacy and Security:** Sensitive data remains on the device rather than being transmitted to cloud servers.
- ✓ **Energy Efficiency:** Optimized hardware such as AI accelerators, GPUs, TPUs, and NPUs enables efficient AI processing.

✚ Basic Architecture of Edge AI:

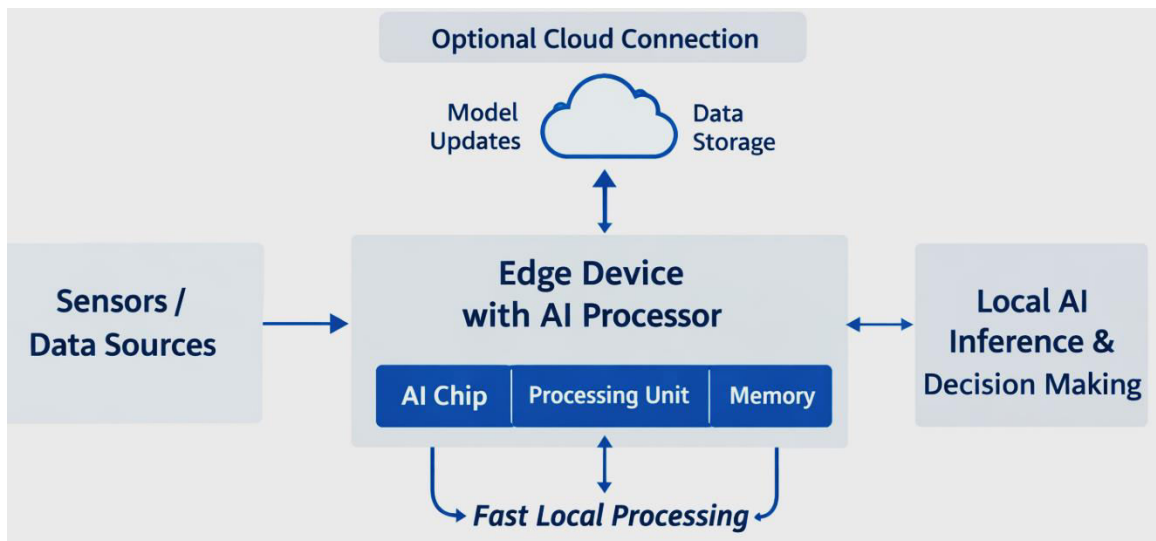


Fig (1): Architecture of Edge AI

- ✓ **Sensors / Data Sources:** Capture data such as images, temperature, voice, or motion.
- ✓ **Edge Device with AI Processor:** Processes data locally using embedded AI models.
- ✓ **Local Decision Making:** The device takes action immediately based on AI analysis.
- ✓ **Optional Cloud Connection:** The cloud may be used for model updates, long-term storage, or large-scale analysis.

❖ Applications of Edge AI in Daily Life Devices:

- ✓ Edge AI is widely used in many smart devices and home appliances to enable real-time decision-making without relying on cloud servers. By processing data locally, these devices become faster, more efficient, and capable of intelligent automation.\
- 1. **Smart Refrigerators:** Edge AI enables smart refrigerators to monitor and manage food items intelligently.
 - ✓ Detects food items using internal cameras and image recognition.
 - ✓ Tracks expiry dates and suggests recipes based on available ingredients.
 - ✓ Adjusts cooling levels automatically depending on stored items.
 - ✓ Sends alerts when food items are running low or expired.
- 2. **Smart LED Bulbs:** Edge AI helps smart lighting systems become energy efficient and automated.
 - ✓ Adjusts brightness and color temperature automatically based on environment.
 - ✓ Detects human presence using motion sensors.
 - ✓ Turns lights ON/OFF automatically to save energy.
 - ✓ Learns user lighting preferences over time.
- 3. **Smart Surveillance Cameras:** Edge AI allows surveillance cameras to analyze video locally and detect events instantly.
 - ✓ Performs face recognition and object detection.
 - ✓ Distinguishes between humans, animals, and vehicles.
 - ✓ Reduces false alarms by intelligent motion detection.
 - ✓ Sends real-time alerts when suspicious activity is detected.
- 4. **Smart Microwave Ovens:** Edge AI improves the cooking experience by automatically adjusting cooking parameters.
 - ✓ Recognizes food type and weight automatically.
 - ✓ Adjusts cooking time and temperature automatically.
 - ✓ Prevents overcooking or undercooking of food.
 - ✓ Suggests optimal cooking settings for different meals.
- 5. **Smart Cars and Scooters:** Edge AI plays a major role in modern intelligent transportation systems.
 - ✓ Enables driver assistance systems such as lane detection and collision warning.
 - ✓ Detects pedestrians, vehicles, and obstacles in real time.
 - ✓ Supports smart navigation and traffic analysis.
 - ✓ Improves vehicle safety and autonomous driving features.

❖ Edge AI in Smart Appliances:

- ✓ Edge AI in smart appliances refers to integrating Artificial Intelligence directly into household devices so they can process data locally and make intelligent decisions without relying heavily on cloud servers.
- ✓ These appliances use sensors, embedded processors, and AI models to analyze data and respond automatically in real time. This enables faster response, reduced internet dependency, and improved privacy, while also helping appliances learn user preferences and optimize their operation.

🌈 Key Features of Edge AI in Smart Appliances:

- ✓ **Local Data Processing:** Smart appliances process data directly on the device using embedded AI chips or microcontrollers.

- ✓ **Real-Time Decision Making:** Appliances can immediately respond to sensor data without waiting for cloud communication.
- ✓ **Energy Efficiency:** AI algorithms optimize power usage by automatically adjusting device settings.
- ✓ **Personalized User Experience:** Smart appliances learn user habits and preferences to improve comfort and convenience.
- ✓ **Improved Reliability:** Devices can continue functioning even when internet connectivity is limited.

✚ **Examples of Smart Appliances Using Edge AI:**

- ✓ Smart refrigerators that monitor food items and adjust cooling automatically.
- ✓ Smart microwave ovens that determine optimal cooking time for different foods.
- ✓ Smart air conditioners that regulate temperature based on room occupancy.
- ✓ Smart washing machines that automatically select washing cycles based on load type.

MH\$MH\$MH\$MH